



[Home](#) > [Statistics](#) > [Detailed methodology information](#) > [Information papers](#)
> [Future directions for modelled regional labour force estimates using administrative data](#)

Future directions for modelled regional labour force estimates using administrative data

New methods and plans for modelled regional labour market estimates by age and sex and below the SA4 level

Released 28/06/2024

Source: [Labour Force, Australia, Detailed, May 2024](#)
[Labour Force, Australia, May 2024](#)

On this page

- [Background](#)
- [Adding age and sex to SA4 estimates](#)
- [Producing new SA3 level estimates](#)
- [Release plans and timing](#)

Background

The Australian Bureau of Statistics (ABS) [commenced releasing monthly modelled regional estimates for Statistical Area Level 4s \(/media-centre/media-releases/better-monthly-regional-labour-market-data-now-available\)](#) (SA4s) in [Labour Force, Australia, Detailed \(https://www.abs.gov.au/statistics/labour/employment-and-unemployment/labour-force-australia-detailed/mar-2024\)](#) in May 2024.

These new estimates are produced using a Rao-Yu time series model, which draws on the Labour Force Survey direct estimates along with Single Touch Payroll (STP) data from the Australian Taxation Office and JobSeeker and Youth Allowance recipients data from the Department of Social Services. Further information on this method can be found in:

- [Improving SA4 level estimates from the Labour Force Survey using administrative data models \(https://www.abs.gov.au/statistics/detailed-methodology-information/information-papers/improving-sa4-level-estimates-labour-force-survey-using-administrative-data-models\)](#)
- [Further refinements to modelled SA4 level Labour Force estimates using administrative data \(https://www.abs.gov.au/statistics/detailed-methodology-information/information-papers/further-refinements-modelled-sa4-level-labour-force-estimates-using-administrative-data\)](#)
- [A Rao-Yu model for small area estimation of labour force statistics \(https://www.abs.gov.au/statistics/research/rao-yu-model-small-area-estimation-labour-force-statistics\)](#)

The ABS has recently been exploring how to extend these new methods to produce additional regional labour force data. This paper outlines the methods ABS are exploring to model:

- age and sex information for the new monthly SA4 estimates, and
- labour force estimates at the SA3 level.

Feedback

The paper also provides some preliminary estimates from early versions of these models.

Adding age and sex to SA4 estimates

Progressing a new method through three development phases

The recently released modelled SA4 level estimates use a Rao-Yu time series model, which draws on Labour Force Survey direct estimates along with administrative data (Single Touch Payroll (STP) data and JobSeeker and Youth Allowance recipients data). The proposed final model for producing sex and age information for SA4s would fit the coefficients of the Rao-Yu models to SA4 and state/age/sex series simultaneously, and then apply those coefficients to the disaggregated administrative data.

The key strength of this approach is that it reflects many local level factors that affect certain age groups or sexes differently (e.g. the COVID pandemic-related lockdowns affected SA4s in different ways and the job losses tended to be higher amongst the younger population, who were also more likely to be in casual employment and/or customer-facing roles).

To ensure the available data and methods will support the level of detail required, a three-phase approach of refining and testing modelling approaches is being used in the development of these new estimates. Each phase adds increased complexity to the model, through utilising more detailed data and more complex data models.

The interim model from the first phase

At this stage, the ABS has completed the first of these three phases, which the initial estimates in this information paper are based on. It is for that reason that the ABS is referring to this as the 'interim model', as it will be finessed further, as the ABS progresses towards the model that will be used to produce these estimates each month.

The interim model, based on the first phase of this work uses a relatively simple approach that takes age and sex employment and population information at the national level and applies this to the modelled labour force status estimates at the SA4 level. For each month, the proportions of each age/sex group that are employed and unemployed is multiplied by the number of people in each SA4 in each age/sex group. This is then calibrated to the SA4 employment and unemployment totals produced by the Rao-Yu model. This is the approach used to produce the indicative estimates provided later in this information paper.

This interim model provides some useful indicative data, however it does have some limitations. The data has only been calibrated in a way that the number of unemployed people in each SA4 adds up to totals for state/territory and Australia. However, the sum of unemployed people in age groups in SA4s in a particular state may not add up to the published totals for the age groups in that state.

Also, if an SA4 has a higher unemployment rate than the national unemployment rate, all age and sex unemployment estimates for the SA4 will be higher than the national equivalents by the same factor. The inherent limitation of the interim model, which the ABS will address in the second and third phases, is that it implicitly assumes that age and sex employment and unemployment patterns are constant across all SA4s within a state.

A more detailed description of this interim method is available below.

Methodology of the interim model

Let a be the age/sex group. Age is one of (15-24, 25-34, 35-44, 55-64, 65+); sex is one of (Male, Female); so a ranges through: ("15-24 Males", ... "65+ Females").

Let S be the state of residence and d be the SA4 of residence.

Let P be population counts; so P_a is the Australia-level total of people in an age/sex group, and $P_{a,d}$ is the SA4 level

total for that age group.

Let \hat{y}_a be the count of employed people, as published from direct survey estimates, by age/sex group at the Australia level. Let $\hat{\theta}_d$ be the modelled estimates of employment counts, as published at the SA4 level.

In the simplest approach, the fraction of employed \hat{f}_a in an age/sex group is estimated as:

$$\hat{f}_a = \frac{\hat{y}_a}{P_a}$$

The estimate at SA4/age/sex level is then:

$$\hat{\theta}_{d,a} = \hat{f}_a P_{a,d} \times \frac{\hat{\theta}_d}{\sum_a \hat{f}_a P_{a,d}}$$

In this equation, the fraction serves to calibrate the fine estimates to the SA4 total.

This calculation is performed identically for unemployment, and for every month where all input data is available.

In the second approach, \hat{f}_a should be replaced with a state-specific equivalent $\hat{f}_{a,S}$ but because the direct survey estimate $\hat{y}_{a,S}$ is excessively volatile, the proposal is to substitute smoother estimates $\hat{\theta}_{a,S}$ from small area estimation with a Rao-Yu model.

$$\hat{f}_{a,S} = \frac{\hat{\theta}_{a,S}}{P_{a,S}}$$

The second and third phases of refining the model

The next and second phase of refining the model will involve incorporating age/sex propensities at the state and territory level (rather than at the national level). While allowing state and age/sex interactions to be modelled, this approach would still not be able to directly reflect local level factors in the data.

The third and final phase of model refinements will involve refining the phase one and two methods described above, and investigating using the age/sex breakdowns at the SA4 level directly available in the administrative data. This would reflect more direct local level age/sex factors in the final model to be used to produce monthly SA4 estimates by age and sex on an ongoing basis.

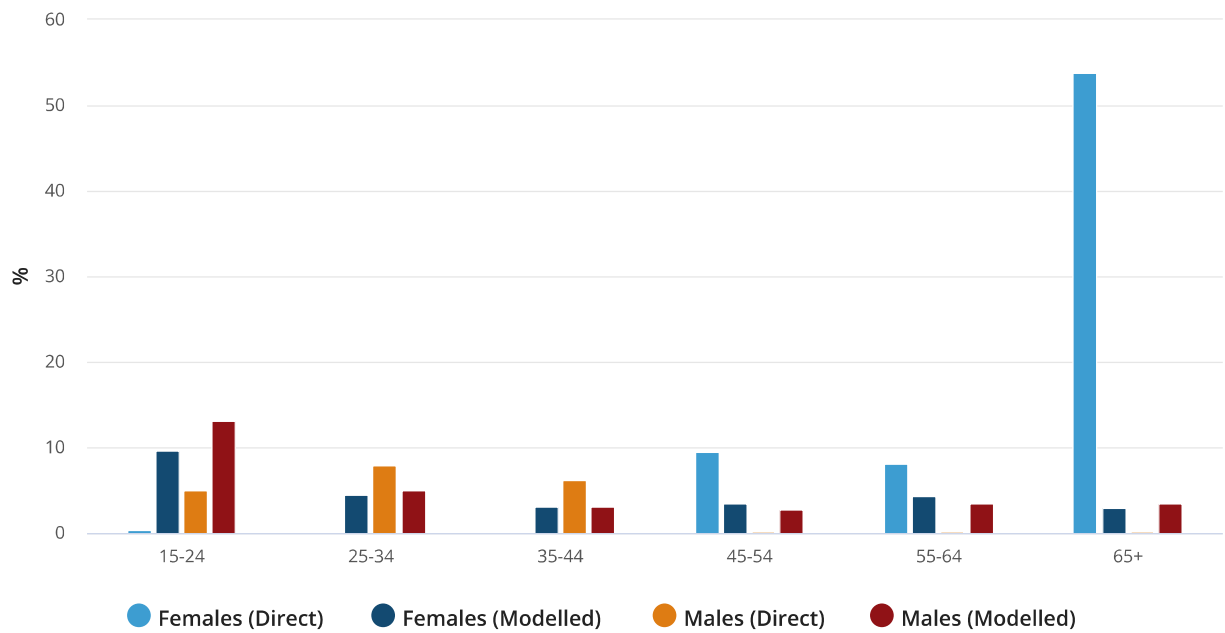
Interim results from the first phase of age and sex modelling for SA4s

Illustrative estimates of labour force status, by age and sex, for SA4s are available in the spreadsheet and graphs below.

Illustrative modelled labour force estimates, by SA4, Age and Sex

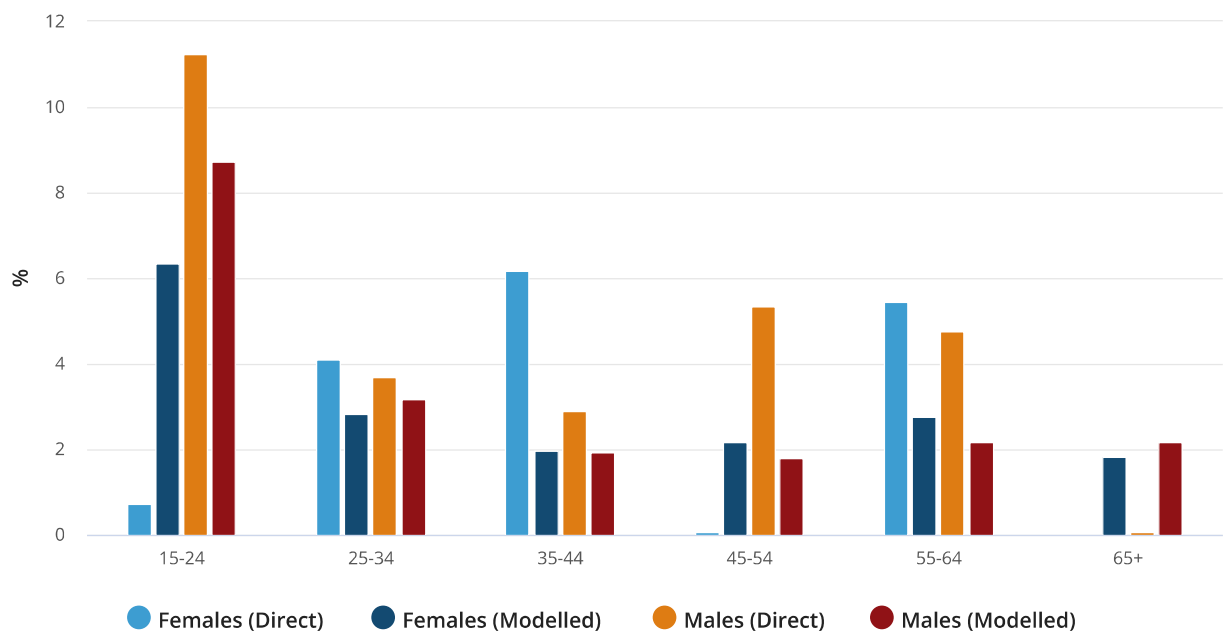
↓ [Download XLSX](#)
[7.81 MB]

Unemployment rate by age and sex(a), Townsville, January 2024



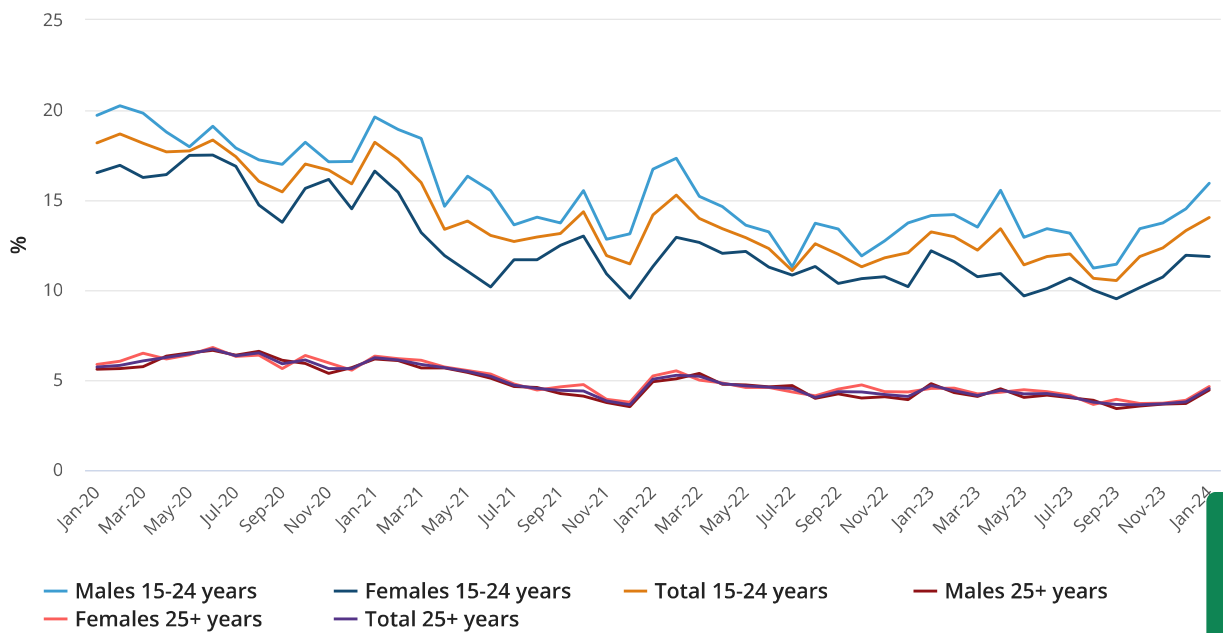
(a) Given the small sample sizes in the direct survey estimates, for some age and sex breakdowns no unemployed people were observed in the sample.

Unemployment rate by age and sex(a), Brisbane - Inner City, January 2024

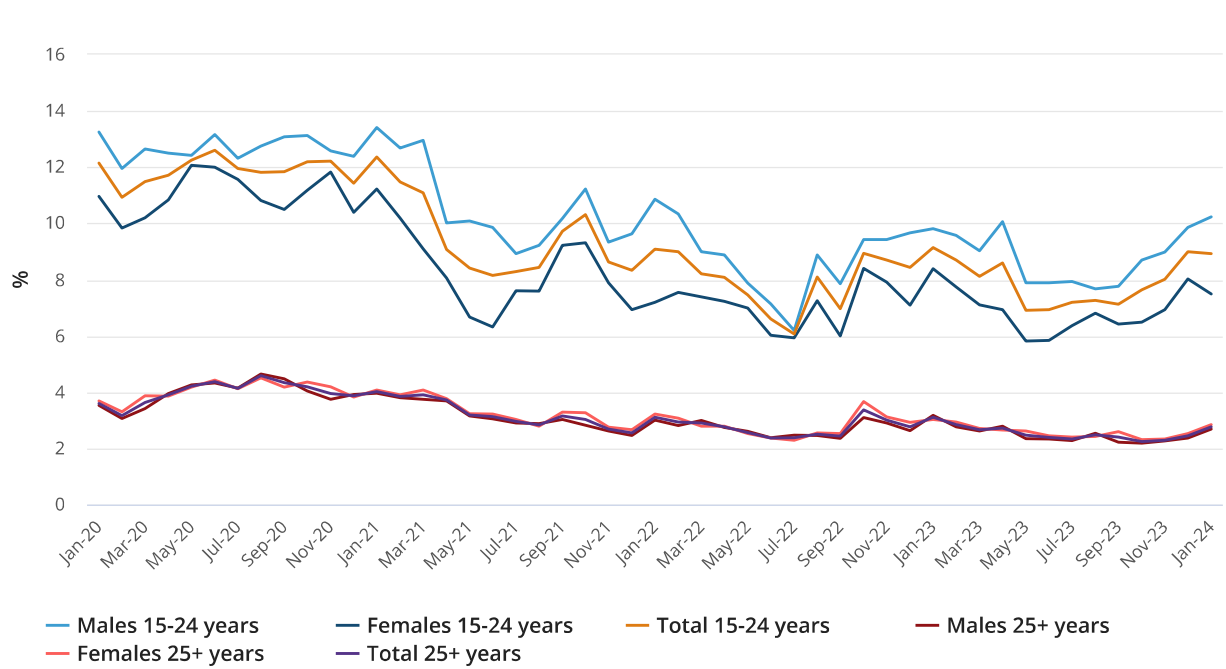


(a) Given the small sample sizes in the direct survey estimates, for some age and sex breakdowns no unemployed people were observed in the sample.

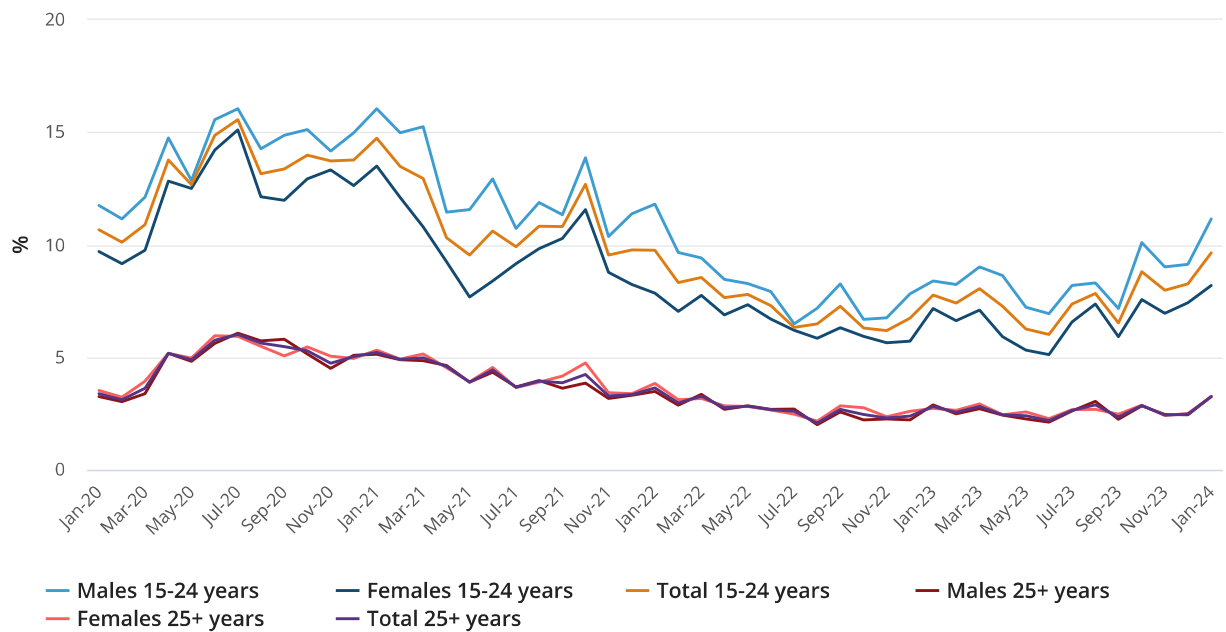
Youth unemployment rate by sex, South Australia - Outback



Youth unemployment rate by sex, Shepparton



Youth unemployment rate by sex, Sydney - Inner City



Producing new SA3 level estimates

Proposed method

The ABS has also started investigating methods to model labour force estimates below the SA4 level, with an initial focus on producing quarterly labour force statistics at the SA3 level.

A similar model to the Rao-Yu model used for the SA4 estimates is used as the foundational model for SA3 level estimates. However, the SA3 model is also fit within a Hierarchical Bayesian framework.

The key differences between the SA4 and SA3 approach are the algorithms used to estimate model parameters and the use of prior distributions for model parameters. Use of the Hierarchical Bayes model for SA3 labour force statistics allows the ABS to fit a wider range of models to the data and explore various extensions of the Rao-Yu model.

Both the Rao-Yu and Hierarchical Bayes models improve small area estimates by borrowing strength across areas, time periods or other variables. Similar to the SA4 model, Single Touch Payroll (STP) data from the Australian Taxation Office and JobSeeker and Youth Allowance recipients data from the Department of Social Services are used.

Importantly, sample sizes in SA3s are significantly smaller than in SA4s, with many observing no unemployed people in the Labour Force Survey sample. It is for that reason that no direct survey estimates have been produced from the Labour Force Survey, with SA4 the lowest level of geographic breakdowns in current Labour Force Survey outputs.

As such, some SA3s are not currently able to be effectively modelled from available data. The illustrative data presented below provides estimates for just over 60% of SA3 areas. Further refinements to the method will aim to produce data for more SA3's.

Although the model is fit using monthly data, quarterly estimates have been derived by averaging over the relevant months.

This modelling strategy takes much inspiration from [Boonstra and Brakel's Hierarchical Bayes models from Statistics Netherlands](https://cris.maastrichtuniversity.nl/en/publications/multilevel-time-series-models-for-small-area-estimation-at-differ) (<https://cris.maastrichtuniversity.nl/en/publications/multilevel-time-series-models-for-small-area-estimation-at-differ>).



Detailed methodology

Modelling

Let x_{dt} be a vector of covariates for SA3 d at month t , and let y_{dt} denote the direct estimate of proportion of employed or unemployed people. Our model structure can be described by:

$$y = x'\beta + v + u + e$$

where:

- β is a vector of state-specific fixed effect coefficients. Additionally, random effects are included which permit the coefficient associated with STP to vary between SA3s, and the coefficient associated with benefit recipients data to vary between SA3s across time
- $v \sim N(0, \sigma_v^2)$ is an SA3 level random intercept
- u is an SA3-by-month random effect modelled as a first order random walk: $u_{dt} = u_{d,t-1} + \epsilon_{dt}$ and $\epsilon_{dt} \sim N(0, \sigma_\epsilon^2)$
- $e \sim N(0, \Psi)$ is a vector of sampling errors with covariance matrix Ψ . We assume the covariance matrix is a block-diagonal matrix of the form $\Psi = \oplus_d \lambda_d \Psi_d$. Here Ψ_d is the covariance matrix block for a particular SA3 and λ_d is a scaling parameter to be estimated

Explanatory variables are derived from monthly STP and DSS counts for each SA3. These are divided by estimates of the civilian population aged 15 and over to represent the proportion of employed and unemployed people. SA3 level

population benchmarks are only available at a quarterly frequency so linear imputation was used to derive monthly values. In addition, a rest-of-state covariate is derived from the GCCSA classification. State-specific DSS proportion and rest-of-state status are used as fixed effects in both the employment and unemployment models, while STP proportion is only used as a fixed effect in the employment model.

Bayesian modelling requires the specification of prior distributions for the model parameters. Weakly informative normal priors have been used for the coefficients β and inverse-chi-square priors with one degree of freedom are used for the variance parameters σ_v^2 and σ_e^2 and the scalars λ_d .

As noted above, the SA3-by-month random effect is modelled using a first order random walk. In contrast, the ABS Rao-Yu model for SA4s uses a first order autoregressive process of the form $u_{dt} = u_{d,t-1} - 1 + \epsilon_{dt}$ and $\epsilon_{dt} \sim N(0, \sigma_e^2)$, where ρ is a parameter to be estimated. While mcmcscs allows us to fit models with a known AR(1) parameter ρ , we are not currently able to use it to estimate ρ .

The covariance matrix is block-diagonal with each block representing a particular SA3. Direct variance estimates for SA3s are volatile due to sampling error and smoothing is required to construct a covariance matrix. Direct variance estimates are first smoothed across SA3s for each month. Based on previous work of [Zhang and Honchar \(2017\)](https://www.abs.gov.au/ausstats/abs@.nsf/mf/1351.0.55.159) (<https://www.abs.gov.au/ausstats/abs@.nsf/mf/1351.0.55.159>), each block is modelled as a second order autoregressive process. Smoothed direct estimates are combined with previously derived state-specific AR(2) parameters to construct the blocks of the covariance matrix.

Deriving statistics and uncertainty estimates from posterior samples

The main output of the model fitting process is a set of 'posterior samples' which represent the probability densities of the various model parameters, given the observed data. A benefit of the Bayesian modelling approach is that parameter estimates, fitted values, and uncertainty estimates are derived as summary statistics of the posterior samples. Estimates of unemployment rate and other statistics which can be expressed in terms of model parameters and observed data are also derived from summary statistics of transformed posterior samples.



Working within the Hierarchical Bayes framework allows various extensions of the SA3 model described above to be explored. Further refinements that will be explored before finalising the model include:

- the introduction of random effects which account for changes in the relationship between STP data and Labour Force Survey data over time
- the use of conditional autoregressive components to model random effects which are spatially correlated at the local level
- the use of non-normally distributed random effects

In the future, the ABS also intends to explore the possibility of either extending the model, or developing a new model, to produce labour force estimates at the SA2 level.

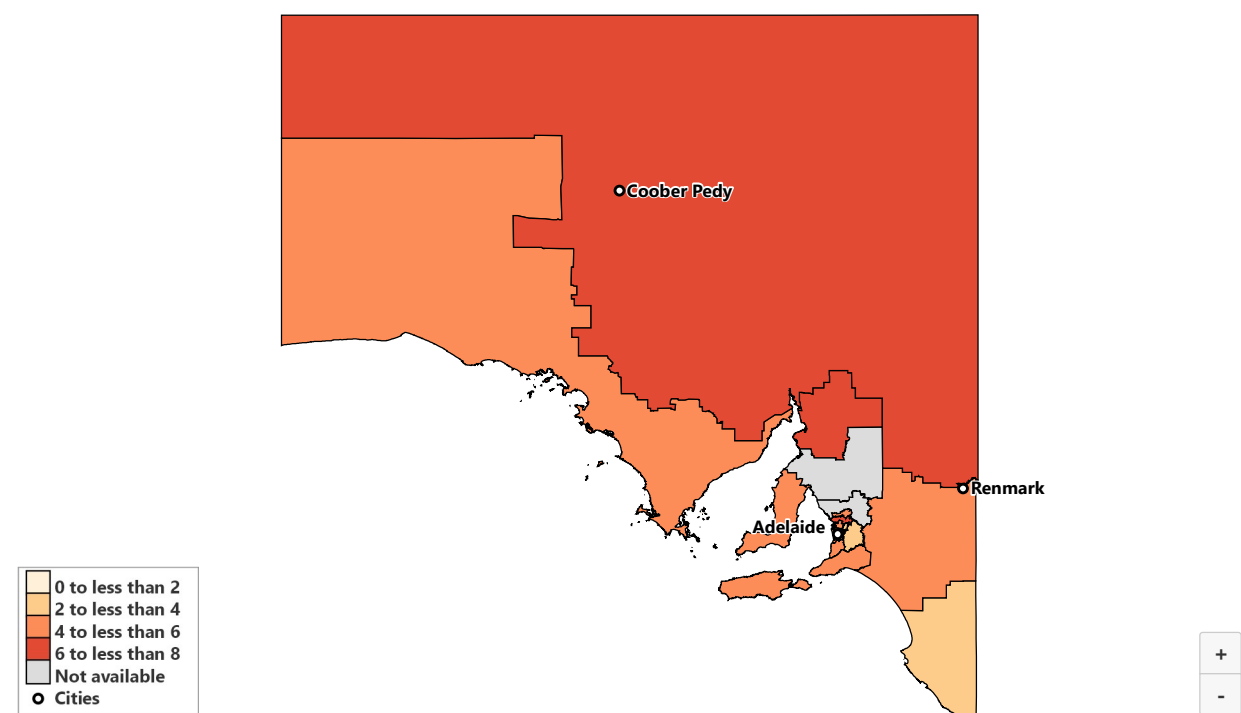
Interim results from the first phase of modelling for SA3s

Illustrative estimates of labour force status, for selected SA3s are available in the spreadsheet below, map and graphs below.

Illustrative estimates by labour force status, by selected SA3s

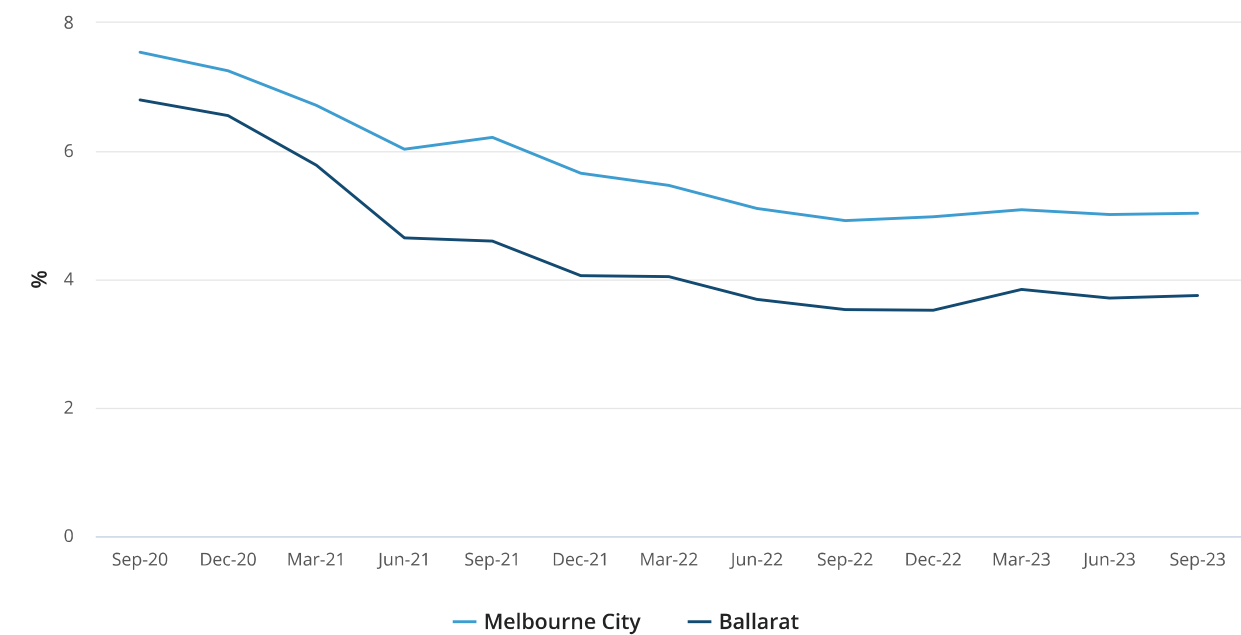
[↓ Download XLSX](#)
[320.58 KB]

Unemployment rates for South Australia by SA3(a), September Quarter 2023

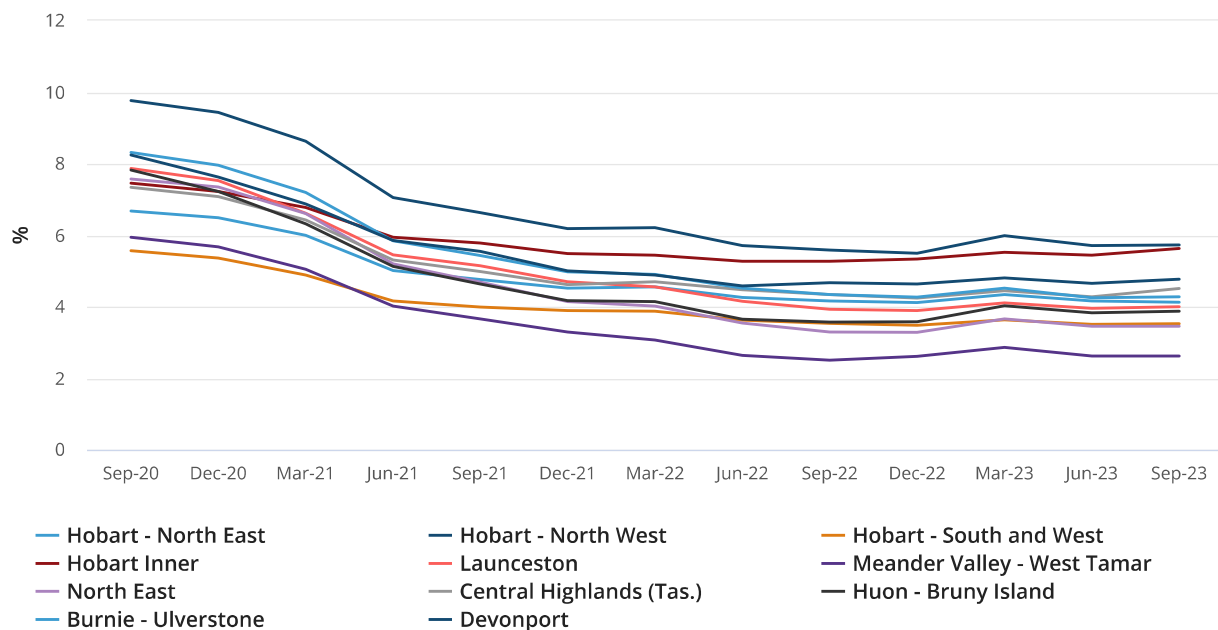


(a) Illustrative estimates are only available for a selection of SA3s.

Unemployment rate, Melbourne City and Ballarat (Original), Sep 2020 - Sep 2023



Unemployment rate, Tasmania, Selected SA3s (Original), Sep 2020 - Sep 2023



Release plans and timing

The ABS is aiming to commence releasing SA4 level estimates of labour force status by age and sex by the end of 2024. The data will be released in [Labour Force, Australia, Detailed \(https://www.abs.gov.au/statistics/labour/employment-and-unemployment/labour-force-australia-detailed/latest-release\)](https://www.abs.gov.au/statistics/labour/employment-and-unemployment/labour-force-australia-detailed/latest-release), and will be added to the 'MRM1 - Modelled estimates of labour force status, by SA4 ASGS)' spreadsheet.

After an appropriate transition period, the spreadsheets containing the direct survey estimates that the modelled estimates have replaced (i.e. Table 16, Table 16b and RM1), will no longer be published in [Labour Force, Australia, Detailed \(https://www.abs.gov.au/statistics/labour/employment-and-unemployment/labour-force-australia-detailed/latest-release\)](https://www.abs.gov.au/statistics/labour/employment-and-unemployment/labour-force-australia-detailed/latest-release). However, the direct survey data will continue to be available from the [Longitudinal Labour Force \(/statistics/microdata-tablebuilder/available-microdata-tablebuilder/longitudinal-labour-force-australia\)](https://www.abs.gov.au/statistics/microdata-tablebuilder/available-microdata-tablebuilder/longitudinal-labour-force-australia) which is available in ABS DataLab, and from a future Labour Force Survey TableBuilder product.

In line with this change, ABS also plans to retire the other regional Labour Force Survey products derived from the direct survey estimates (i.e. Table 16c, RM3, RQ1 and RQ2) in the detailed release. While these data are unlikely to be replaced in the foreseeable future by modelled equivalents, the detailed underlying data will also continue to be available from the Longitudinal Labour Force and future Labour Force Survey TableBuilder product.

The ABS expects to be able to commence releasing estimates at the SA3 level by the end of 2025 as an additional spreadsheet in the detailed Labour Force release. Further details on this will be provided closer to the release.